JOURNAL OF COMPLEXITY **18**, 394–414 (2002)
doi:10.1006/jcom.2001.0627

Linear vs Standard Information for Scalar Stochastic Differential Equations

Norbert Hofmann

Fachbereich Mathematik, Johann Wolfgang Goethe-Universität, Postfach 11 19 32, 60054 Frankfurt am Main, Germany
E-mail: hofmann@math.uni-frankfurt.de

and

Thomas Müller-Gronbach and Klaus Ritter

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany

Received April 2, 2001; revised August 24, 2001; accepted August 24, 2001; published online March 7, 2002

We study pathwise approximation of scalar sde's with respect to the mean squared L_2-error. We compare the power of linear and standard information about the driving Brownian motion. It turns out that asymptotically the corresponding minimal errors differ only by the factor $\sqrt{6}/\pi$.

1. INTRODUCTION

The Ito–Taylor expansion yields a family of numerical methods for strong approximation of stochastic differential equations. This family includes in particular the Euler, Milstein, and Wagner–Platen schemes. With increasing length of the expansion, Ito–Taylor methods need to evaluate more and more complicated functionals of the driving Brownian motion W.

In this paper we study the scalar case; i.e., W as well as the solution X of

$$dX(t) = a(t, X(t))\, dt + \sigma(t, X(t))\, dW(t), \qquad t \in [0, 1], \tag{1}$$

0885-064X/02 $35.00

is a real-valued process. For the Euler and Milstein schemes it then suffices to evaluate W at finitely many points; i.e., only Dirac functionals are applied to the trajectories of W. The Wagner–Platen scheme additionally needs (Riemann) integrals of W over subintervals of $[0, 1]$; i.e., certain bounded linear functionals are applied to the trajectories of W. We investigate whether the whole class of bounded linear functionals is actually more powerful for strong approximation than the subclass of Dirac functionals.

A strong approximation yields a process $\bar{X}$ whose paths are close to the respective paths of the solution X of (1). We study the pathwise distance between X and $\bar{X}$ globally on $[0, 1]$ in the L_2-norm $\|\cdot\|_2$, and the error of $\bar{X}$ is defined by

$$e(\bar{X}) = (E(\|X - \bar{X}\|_2^2))^{1/2}.$$

As a rough measure of cost we use the average number $n(\bar{X})$ of functionals from a class Λ that are applied to the Brownian motion. The functionals as well as their total number may be chosen adaptively.

The power of a class Λ is now expressed by the sequence of minimal errors

$$e^{**}(N, \Lambda) = \inf\{e(\bar{X}) : n(\bar{X}) = N, \bar{X} \text{ uses functionals from } \Lambda\}.$$

By definition, $e^{**}(N, \Lambda)$ is the smallest error that can be obtained by any method that uses N functionals from Λ on the average.

Here we focus on linear information about W, i.e.,

$$\Lambda = \Lambda^{\mathrm{lin}}, \qquad \text{the class of all bounded linear functionals on } C([0, 1]),$$

and we show that

$$\lim_{N \to \infty} N^{1/2} \cdot e^{**}(N, \Lambda^{\mathrm{lin}}) = 1/\pi \cdot E\left(\int_0^1 |\sigma(t, X(t))| \, dt\right).$$

To prove the asymptotic upper bound, we construct a new method for strong approximation, which uses Milstein steps and, locally, Karhunen–Loève expansions with an adaptively chosen number of terms.

For comparison, consider the case of standard information about W, i.e.,

$$\Lambda = \Lambda^{\mathrm{std}}, \qquad \text{the class of all Dirac functionals on } C([0, 1]),$$

which was already studied in Hofmann *et al.* (2001). It turns out that

$$\lim_{N\to\infty} \frac{e^{**}(N, \Lambda^{\text{lin}})}{e^{**}(N, \Lambda^{\text{std}})} = \frac{\sqrt{6}}{\pi} \simeq 0.78; \tag{2}$$

i.e., selecting functionals from $\Lambda^{\text{lin}} \setminus \Lambda^{\text{std}}$ does not help much, asymptotically. Moreover, the same result holds if we restrict considerations to methods with fixed cardinality or methods that apply a fixed set of functionals to all trajectories of W.

The methods of proof are similar for $\Lambda = \Lambda^{\text{lin}}$ and $\Lambda = \Lambda^{\text{std}}$, namely, reducing the strong approximation problem for a stochastic differential equation to an L_2-approximation problem for randomly weighted Brownian bridges. Asymptotically optimal methods for strong approximation are easy to implement in both cases, and numerical experiments are presented in Hofmann *et al.* (2001) for the class Λ^{std}.

The relation (2) was already known for the trivial equation $dX(t) = dW(t)$; see Lee (1986) and Papageorgiou and Wasilkowski (1990). This particular equation defines a linear problem with a Gaussian measure, which is given by the embedding of $C([0, 1])$ into $L_2([0, 1])$ and the Wiener measure on $C([0, 1])$. For many linear problems of this type it is known that the class Λ^{std} is almost as powerful as the class Λ^{lin}; see, e.g., Traub *et al.* (1988), Ritter (2000), and Wasilkowski and Woźniakowski (2001) for results and references.

Frequently, $\bar{X}$ and X are only compared at the single point $t = 1$, and the corresponding error is defined by

$$d(\bar{X}) = (E(X(1) - \bar{X}(1))^2)^{1/2}.$$

Then upper bounds of order n^{-1} and $n^{-3/2}$ hold for the Milstein scheme and the Wagner–Platen scheme, respectively, both using a fixed step-size $1/n$. See Milstein (1974) and Wagner and Platen (1978). On the other hand, for the corresponding minimal errors $d^{**}(N, \Lambda^{\text{std}})$ a lower bound of order $1/n$ holds for most equations, see Müller–Gronbach (2001). Thus bounded linear functionals are much superior to Dirac functionals if the error at $t = 1$ is studied.

2. PROBLEM FORMULATION AND RESULTS

2.1. *Assumptions.* Throughout this paper we assume that the drift and diffusion coefficients

$$a, \sigma : [0, 1] \times \mathbb{R} \to \mathbb{R}$$

and the initial value $X(0)$ have the following properties.

(A) Both, a and σ are differentiable with respect to the state variable. Moreover, there sexists a constant $K>0$ such that $f=a$ and $f=\sigma$ satisfy

$$
\begin{aligned}
|f(t,x)-f(t,y)| &\leqslant K\cdot|x-y|,\\
|f(s,x)-f(t,x)| &\leqslant K\cdot(1+|x|)\cdot|s-t|,\\
|f^{(0,1)}(t,x)-f^{(0,1)}(t,y)| &\leqslant K\cdot|x-y|
\end{aligned}
$$

for all $s, t\in[0,1]$ and $x, y\in\mathbb{R}$

(B) The initial value $X(0)$ is independent of W and

$$E(X(0))^4<\infty.$$

Note that (A) yields the linear growth condition

$$|f(t,x)|\leqslant c\cdot(1+|x|)$$

with a suitable constant $c>0$ and the boundedness of $f^{(0,1)}$. Given (A) and (B), a pathwise unique strong solution of the Eq. (1) with initial value $X(0)$ exists. In particular,

$$\sup_{t\in[0,1]} E(X(t))^4<\infty. \tag{3}$$

2.2. *The Class of Methods.* We study methods for pathwise approximation that are based on

(1) complete knowledge of the drift and diffusion coefficients,

(2) the realization of the initial value,

(3) adaptive (sequential) selection of a finite number of bounded linear functionals that are applied to the trajectory of the Brownian motion.

Let Λ^{lin} denote the space of bounded linear functionals on $C([0,1])$. Fix a and σ, and consider the corresponding equation (1). Formally, a general method is then defined by mappings

$$
\begin{aligned}
\psi_k &: \mathbb{R}^k\to\Lambda^{\mathrm{lin}},\\
\chi_k &: \mathbb{R}^{k+1}\to\{\mathrm{STOP},\mathrm{GO}\},\\
\phi_k &: \mathbb{R}^{k+1}\to L_2([0,1])
\end{aligned}
$$

for $k \in \mathbb{N}$. The realization x of the initial value determines the first functional $\psi_1(x)$ that is applied to the trajectory w of the Brownian motion. After k steps we have observed the data

$$\Psi_k(x, w) = (x, y_1, \ldots, y_k),$$

where

$$y_1 = \psi_1(x)(w), \ldots, y_k = \psi_k(x, y_1, \ldots, y_{k-1})(w).$$

A decision to stop or to apply another functional to w is made after each step, and the total number of functionals that are applied to w is given by

$$\nu(x, w) = \min \{k \in \mathbb{N} : \chi_k(\Psi_k(x, w)) = \text{STOP}\}.$$

If $\nu(x, w) < \infty$ then the data

$$\Psi(x, w) = \Psi_{\nu(x, w)}(x, w)$$

are used to construct the approximation $\psi_{\nu(x, w)}(\Psi(x, w))$.

We only assume Borel measurability of the mappings χ_k, ϕ_k, and $\psi_k(\cdot)(w)$ for every $w \in C([0, 1])$. This ensures that all mappings Ψ_k are Borel measurable on $\mathbb{R} \times C([0, 1])$.

Obviously only the case $\nu(X(0), W) < \infty$ with probability one is of practical interest. Then we end up with the method

$$\bar{X} = \phi_{\nu(X(0), W)}(\Psi(X(0), W)).$$

We relate the error $e(\bar{X})$ to the expected number

$$n(\bar{X}) = E(\nu(X(0), W))$$

of functionals used by $\bar{X}$.

2.3. *Minimal Errors.* Let χ^{**} denote the class of all methods of the above form, and put

$$\chi_N^{**} = \{\bar{X} \in \chi^{**} : \lceil n(\bar{X}) \rceil = N\}$$

for $N \in \mathbb{N}$. The quantity

$$e^{**}(N) = \inf \{e(\bar{X}) : \bar{X} \in \chi_N^{**}\}$$

is the minimal error that can be obtained by methods that use N sequentially chosen functionals on the average.

As a subclass $\chi^* \subset \chi^{**}$ we consider all methods that apply the same number of functionals to all trajectories of w. Formally this means that the mappings χ_k are constant and $v = \min\{k \in \mathbb{N} : \chi_k = \text{STOP}\}$. We put

$$\chi_N^* = \{\bar{X} \in \chi^* : n(\bar{X}) = N\}$$

as well as

$$e^*(N) = \inf\{e(\bar{X}) : \bar{X} \in \chi_N^*\}.$$

The subclass $\chi \subset \chi^*$ consists of all methods that use the same functionals for every trajectory. Formally the mappings ψ_k and χ_k are constant, such that $\Psi(x, w) = (x, \psi_1(w), ..., \psi_v(w))$. We put

$$\chi_N = \{\bar{X} \in \chi : n(\bar{X}) = N\}$$

as well as

$$e(N) = \inf\{e(\bar{X}) : \bar{X} \in \chi_N\}.$$

2.4. *Results and Remarks.* To every Eq. (1) we associate the constants

$$C^{**} = E\left(\int_0^1 |\sigma(t, X(t))|\, dt\right),$$

$$C^* = \left(E\left(\left(\int_0^1 |\sigma(t, X(t))|\, dt\right)^2\right)\right)^{1/2},$$

$$C = \int_0^1 (E(\sigma^2(t, X(t))))^{1/2}\, dt.$$

THEOREM 1. *The minimal errors satisfy*

(i) $\lim_{N\to\infty} N^{1/2} \cdot e^{**}(N) = C^{**}/\pi,$

(ii) $\lim_{N\to\infty} N^{1/2} \cdot e^*(N) = C^*/\pi,$

(iii) $\lim_{N\to\infty} N^{1/2} \cdot e(N) = C/\pi$

for every Eq. (1).

Remark 1. The upper bounds in Theorem 1 are proven constructively. We present asymptotically optimal methods in the three classes χ^{**}, χ^{*}, and χ, all of which are based on the following principle. The Milstein scheme is used to obtain an approximation $\check{X}$ at a coarse equidistant discretization

$$t_\ell = \ell/k, \qquad \ell = 0, ..., k. \tag{4}$$

On every subinterval $]\, t_\ell, t_{\ell+1}]$ we consider the linear interpolation $\tilde{W}$ of $W(t_\ell)$ and $W(t_{\ell+1})$ as well as a Karhunen–Loève approximation R of the Brownian bridge $W - \tilde{W}$. At $t \in \,]\, t_\ell, t_{\ell+1}]$, the solution $X(t)$ is approximated by

$$\check{X}(t_\ell) + a(t_\ell, \check{X}(t_\ell)) \cdot (t - t_\ell) + \sigma(t_\ell, \check{X}(t_\ell)) \cdot (\tilde{W}(t) - W(t_\ell) + R(t)).$$

The number of terms in R depends on $\sigma(t_\ell, \check{X}(t_\ell))$ and $\sigma(t_0, \check{X}(t_0)), ..., \sigma(t_{k-1}, \check{X}(t_{k-1}))$ for the asymptotically optimal methods in χ^{**} and χ^{*}, respectively. See Section 3.4 for details.

We stress that these asymptotically optimal methods do not need complete knowledge of the drift and diffusion coefficients. Instead, it suffices to evaluate a, σ, and $\sigma^{(0,1)}$ at finitely many points. Moreover, the number of these evaluations is asymptotically negligible as compared to the number of functionals that are applied to the Brownian motion.

Remark 2. In the literature, different classes Λ of permissible functionals are studied for strong approximation. For every such class we obtain classes of methods $\mathbb{X}_\Lambda^{**}$, $\mathbb{X}_\Lambda^{*}$, and $\mathbb{X}_\Lambda$ as well as corresponding minimal errors $e^{**}(N, \Lambda)$, $e^{*}(N, \Lambda)$, and $e(N, \Lambda)$ in a canonical way. For instance, $\mathbb{X}_{\Lambda^{\text{lin}}}^{**} = \chi^{**}$.

The class $\Lambda = \Lambda^{\text{std}}$ of Dirac functionals, which is used in many cases, yields methods for strong approximation that are based on observation of the Brownian motion W at adaptively chosen points. The corresponding minimal errors are determined in Hofmann *et al.* (2001), namely,

(i) $\lim_{N \to \infty} N^{1/2} \cdot e^{**}(N, \Lambda^{\text{std}}) = C^{**}/\sqrt{6}$,

(ii) $\lim_{N \to \infty} N^{1/2} \cdot e^{*}(N, \Lambda^{\text{std}}) = C^{*}/\sqrt{6}$,

(iii) $\lim_{N \to \infty} N^{1/2} \cdot e(N, \Lambda^{\text{std}}) = C/\sqrt{6}$.

Remark 3. Comparing Theorem 1 and Remark 2 we see that the order of the minimal errors is $1/2$ in all six cases. The asymptotic constants consist of two factors,

$$\gamma \in \{C^{**}, C^{*}, C\} \qquad \text{and} \qquad \kappa(\Lambda) \in \{1/\pi, 1/\sqrt{6}\}.$$

The class Λ only effects the constants via $\kappa(\Lambda)$. The other term γ depends on "the equation," i.e., drift and diffusion coefficients and distribution of the initial value, as well as on the "structure" of the approximating methods, i.e., varying or fixed cardinality ν, or path-independent selection of functionals.

Hence, by using methods that may apply arbitrary bounded linear functionals instead of Dirac functionals to the trajectories of the Brownian motion we can only achieve a small improvement on the level of asymptotic constants.

The constants C^{**}, C^{*}, and C are further investigated in Hofmann *et al.* (2001), and it turns out that huge differences may occur between these constants.

Remark 4. The stochastic Ito–Taylor expansion is the key to a family of methods that are based on multiple stochastic and deterministic integrals. The corresponding classes $\Lambda = \Lambda(\mathscr{A})$ of permissible functionals are defined via so-called hierarchical sets $\mathscr{A}$, see Kloeden and Platen (1995). Theorem 1 and Remark 2 suggest the following conjecture. The minimal errors satisfy

$$\lim_{N \to \infty} N^{1/2} \cdot \varepsilon(N, \Lambda(\mathscr{A})) = \gamma \cdot \kappa(\Lambda(\mathscr{A})),$$

where $\varepsilon \in \{e^{**}, e^{*}, e\}$ and $\gamma \in \{C^{**}, C^{*}, C\}$, accordingly.

2.5. *The Equidistant Wagner–Platen Scheme.* A primary example of a method that uses functionals from $\Lambda = \Lambda^{\mathrm{lin}}$ is the Wagner–Platen scheme; see Wagner and Platen (1978). We use the equidistant discretization (4) with $k = n$ and denote the corresponding Wagner–Platen method by $\hat{X}_n^{\mathrm{equi}}$. Thus $\hat{X}_n^{\mathrm{equi}}(0) = X(0)$ and

$$\hat{X}_n^{\mathrm{equi}}(t_{\ell+1}) = \hat{X}_n^{\mathrm{equi}}(t_\ell) + \sum_{j=1}^{7} g_j(t_\ell, \hat{X}_n^{\mathrm{equi}}(t_\ell)) \cdot Z_{j,\ell}^{(n)}(t_{\ell+1})$$

for $\ell = 0, ..., n-1$, where

$$\begin{aligned}
g_1 &= a & Z_{1,\ell}^{(n)}(t) &= t - t_\ell, \\
g_2 &= \sigma, & Z_{2,\ell}^{(n)}(t) &= W(t) - W(t_\ell), \\
g_3 &= 1/2 \cdot \sigma\sigma^{(0,1)}, & Z_{3,\ell}^{(n)}(t) &= (W(t) - W(t_\ell))^2 - (t - t_\ell), \\
g_4 &= \sigma^{(1,0)} + a\sigma^{(0,1)} - 1/2 \cdot \sigma(\sigma^{(0,1)})^2 & Z_{4,\ell}^{(n)}(t) &= (t - t_\ell) \cdot (W(t) - W(t_\ell)), \\
g_5 &= 1/6 \cdot (\sigma(\sigma^{(0,1)})^2 + \sigma^2\sigma^{(0,2)}), & Z_{5,\ell}^{(n)}(t) &= (W(t) - W(t_\ell))^3, \\
g_6 &= 1/2 \cdot (a^{(1,0)} + aa^{0,1)} + 1/2 \cdot \sigma^2 a^{(0,2)}), & Z_{6,\ell}^{(n)}(t) &= (t - t_\ell)^2, \\
g_7 &= \sigma a^{(0,1)} - \sigma^{(1,0)} - a\sigma^{(0,1)} - 1/2 \cdot \sigma^2\sigma^{(0,2)}, & Z_{7,\ell}^{(n)}(t) &= \int_{t_\ell}^{t} (W(s) - W(t_\ell))\, ds.
\end{aligned}$$

A global approximation on $[0, 1]$ is obtained by piecewise linear interpolation of the data $(t_\ell, \hat{X}_n^{\mathrm{equi}}(t_\ell))$. Due to the presence of the integrals $Z_{7,\ell}^n(t_{\ell+1})$, the Wagner–Platen scheme is not exclusively based on Dirac functionals.

We strengthen the assumption (A). Instead of (A) we now assume

(A′) All partial derivatives appearing in $g_3, \ldots, g_7$ exist and there exists a constant $K > 0$ such that the functions g_j with $j = 1, \ldots, 7$ satisfy

$$|g_j(t, x) - g_j(t, y)| \leqslant K \cdot |x - y|,$$
$$|g_j(s, x) - g_j(t, x)| \leqslant K \cdot (1 + |x|) \cdot |s - t|,$$

for all $s, t \in [0, 1]$ and $x, y \in \mathbb{R}$.

PROPOSITION 1.

$$\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n^{\mathrm{equi}}) = C^{\mathrm{equi}} / \sqrt{6},$$

where

$$C^{\mathrm{equi}} = \left(\int_0^1 E(\sigma^2(t, X(t)))\, dt \right)^{1/2}.$$

Remark 5. In terms of n, the asymptotic performance of the equidistant Wagner–Platen and Milstein schemes coincide, see Proposition 1 and Hofmann *et al.* (2001, Proposition 1). Note, however, that the numbers of functionals are $2n$ and n, respectively. We stress that the Wagner–Platen scheme is much superior to the Milstein scheme for approximation of X at discrete points.

3. PROOFS

The key idea of the proofs is to relate approximation of X using data from W to approximation of (randomly) weighted Brownian bridges. For the latter problem, data from the Brownian bridges themselves is available.

In the sequel, we let c denote unspecified positive constants, which only depend on the constant K from condition (A) as well as on $a(0, 0)$, $\sigma(0, 0)$, and $E(X(0))^4$.

3.1. *The Milstein Process.* Consider the discretization (4). The Milstein process $\check{X}_k$ is defined by $\check{X}_k(0) = X(0)$ and

$$\begin{aligned}\check{X}_k(t) = \check{X}_k(t_\ell) + a(t_\ell, \check{X}_k(t_\ell)) \cdot (t - t_\ell) + \sigma(t_\ell, \check{X}_l(t_\ell)) \cdot (W(t) - W(t_\ell))\\ + 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(t_\ell, \check{X}_k(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell))\end{aligned}$$

for $t \in]\, t_\ell, t_{\ell+1}]$. In particular, $\check{X}_k$ coincides with the Milstein scheme at the points t_ℓ. Observe, however, that $\check{X}_k$ is not a numerical method for approximating the solution X of the stochastic differential equation globally on $[0, 1]$, since it relies on complete knowledge of the trajectories of W.

The Milstein process has the property

$$\sup_{t \in [0,1]} E(\check{X}_k(t))^4 \leqslant c \tag{5}$$

and satisfies the error estimate

$$\sup_{t \in [0,1]} E(X(t) - \check{X}_k(t))^2 \leqslant c/k^2. \tag{6}$$

See, e.g., Hofmann *et al.* (2001, Lemma 11 and Theorem 4).

3.2. *Approximation of Weighted Brownian Bridges.* Fix $k \in \mathbb{N}$, consider the discretization (4), and let the process Y consist of independent Brownian bridges on $[t_\ell, t_{\ell+1}]$ for $\ell = 0, \ldots, k-1$. Consider a weight function ρ with constant values ρ_ℓ on the interior of these subintervals.

Suppose we wish to approximate the trajectories of ρY in the L_2-norm, and that bounded linear functionals from Λ^{lin} may be applied to the trajectories of Y to this end. The formal definition of methods $\overline{\rho Y}$ is analogous to the one from Section 2.2; only the initial value $X(0)$ is now irrelevant. Let $n(\overline{\rho Y})$ denote the expected number of functionals used by $\overline{\rho Y}$.

Suppose, without loss of generality, that $\rho_\ell \neq 0$ for some ℓ. By

$$\mu_1 \geqslant \mu_2 \geqslant \cdots > 0$$

we denote the positive eigenvalues of the covariance kernel of ρY, repeated according to their multiplicity. Moreover, let

$$f_1, f_2, \ldots \in L_2([0, 1])$$

denote a corresponding orthonormal system of eigenfunctions. Let $N \in \mathbb{N}$ and consider a method $\overline{\rho Y}$ with

$$\lceil n(\overline{\rho Y}) \rceil = N. \tag{7}$$

Due to a general result from Wasilkowski (1989) on the optimality of the truncated Karhunen–Loève decomposition,

$$E\,\|\rho Y - \overline{\rho Y}\|_2^2 \geqslant \sum_{i=N+1}^{\infty} \mu_i.$$

Furthermore, this estimate is sharp on the class of all such methods, since

$$E\,\|\rho Y-\overline{\rho Y}\|_2^2=\sum_{i=N+1}^{\infty}\mu_i \tag{8}$$

for

$$\overline{\rho Y}=\sum_{i=1}^{N}\langle\rho Y, f_i\rangle\cdot f_i.$$

Here $\langle\cdot,\cdot\rangle$ denotes the inner product in $L_2([0,1])$.

In the present case, the eigenvalues and eigenfunctions are known explicitly. For $k=1$ and $\rho_0=1$,

$$\mu_i=\frac{1}{\pi^2 i^2}$$

and

$$f_i(t)=\sqrt{2}\cdot\sin(i\pi t),\qquad t\in[0,1].$$

In general, the independence and the scaling property of the Brownian bridges yields the system of nonzero eigenvalues

$$\mu_{i,\ell}=\frac{1}{k^2\pi^2}\cdot\frac{\rho_\ell^2}{i^2} \tag{9}$$

for $i\in\mathbb{N}$ and $\ell\in\{0,\dots,k-1\}$ such that $\rho_\ell\neq 0$. The corresponding eigenfunctions are given by

$$f_{i,\ell}(t)=\sqrt{k}\cdot f_i(k\cdot(t-t_\ell))\cdot 1_{[t_\ell,t_{\ell+1}]}(t). \tag{10}$$

Hence there exist integers $m_\ell\in\mathbb{N}_0$ such that

$$\sum_{\ell=0}^{k-1} m_\ell=N \tag{11}$$

and

$$E\,\|\rho Y-\overline{\rho Y}\|_2^2\geqslant\frac{1}{k^2\pi^2}\cdot\sum_{\ell=0}^{k-1}\frac{\rho_\ell^2}{m_\ell+1} \tag{12}$$

for every method $\overline{\rho Y}$ with (7). This lower bound is asymptotically sharp, and the numbers m_ℓ are obtained by minimizing $\sum_{\ell=0}^{k-1} (\rho_\ell^2 \cdot \sum_{i=m_\ell+1}^{\infty} 1/i^2)$ under the restriction (11).

3.3. *Proof of the Lower Bounds in Theorem* 1. Consider an arbitrary sequence of methods $\bar{X}_N \in \chi_N^{**}$. Take a sequence of positive integers k_N with the properties

$$N^{1/2} = o(k_N) \tag{13}$$

and

$$k_N = o(N). \tag{14}$$

Use the discretization (4) with $k = k_N$, and consider the corresponding Milstein process $\check{X}_{k_N}$. Furthermore, let $\tilde{W}_{k_N}$ denote the piecewise linear interpolation of the Brownian motion W for this discretization. Put

$$\begin{aligned} V_N &= (X(0), W(t_1), \dots W(t_{k_N})), \\ Y_N &= W - \tilde{W}_{k_N}. \end{aligned}$$

Obviously there exist measurable mappings $\rho_{\ell, N} : \mathbb{R}^{k_N+1} \to \mathbb{R}$ such that

$$\rho_{\ell, N}(V_N) = \sigma(t_\ell, \check{X}_{k_N}(t_\ell)), \qquad \ell = 0, \dots, k_N - 1.$$

For $v \in \mathbb{R}^{k_N+1}$ let

$$\rho_N(v) = \sum_{\ell=0}^{k_N} \rho_{\ell, N}(v) \cdot 1_{[t_\ell, t_{\ell+1}[}.$$

We define $X_N^\dagger$ by

$$\begin{aligned} X_N^\dagger(t) = {} & \bar{X}_N(t) - \check{X}_{k_N}(t_\ell) \\ & - a(t_\ell, \check{X}_{k_N}(t_\ell)) \cdot (t - t_\ell) - \sigma(t_\ell, \check{X}_{k_N}(t_\ell)) \cdot (\tilde{W}_{k_N}(t) - W(t_\ell)) \end{aligned} \tag{15}$$

for $t \in [t_\ell, t_{\ell+1}[$.

Instead of $\bar{X}_N$ as a method for pathwise approximation of X, we will study $X_N^\dagger$ as a method for L_2-approximation of weighted Brownian bridges. The Brownian bridges are given by the process Y_N. The weight function is random and given by the process $\rho_N(V_N)$, which is close to $\sigma(\cdot, X(\cdot))$. Clearly, $X_N^\dagger \in \chi^{**}$, and (14) yields

$$\lim_{N \to \infty} n(X_N^\dagger)/n(\bar{X}_N) = 1,$$

since $n(\bar{X}_N) \leqslant n(X_N^\dagger) \leqslant n(\bar{X}_N) + k_N$. Furthermore, $\bar{X}_N \in \chi^*$ implies $X_N^\dagger \in \chi^*$ and $\bar{X}_N \in \chi$ implies $X_N^\dagger \in \chi$. The errors of $\bar{X}_N$ and $X_N^\dagger$ for the respective problems are related as follows.

LEMMA 1.

$$|e(\bar{X}_N) - (E\,\|\rho_N(V_N)\,Y_N - X_N^\dagger\|_2^2)^{1/2}| = o(1/N^{1/2}).$$

Proof. Note that

$$|e(\bar{X}_N) - (E\,\|\check{X}_{k_N} - \bar{X}_N\|_2^2)^{1/2}| = o(1/N^{1/2}),$$

due to (6) and (13).

Let $t \in [t_\ell, t_{\ell+1}[$. Then

$$\begin{aligned}\check{X}_{k_N}(t) - \bar{X}_N(t) = {} & \rho_{\ell,N}(V_N)\,Y_N(t) - X_N^\dagger(t)\\ & + 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(t_\ell, \check{X}_{k_N}(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)).\end{aligned}$$

The boundedness of $\sigma^{(0,1)}$ and the linear growth of σ yield

$$\begin{aligned} & E((\sigma \cdot \sigma^{(0,1)})(t_\ell, \check{X}_{k_N}(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)))^2\\ & \leqslant c \cdot E(\sigma^2(t_\ell, \check{X}_{k_N}(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell))^2)\\ & \leqslant c \cdot (1 + E(\check{X}_{k_N}(t_\ell))^2) \cdot (t - t_\ell)^2\\ & \leqslant c/k_N^2,\end{aligned}$$

where (5) is used for the last estimate. Now the lemma follows from (13). ∎

In the sequel, $v(x, w)$ denotes the total number of functionals that are used by $\bar{X}_N$ for the realization x of the initial value and the trajectory w of the Brownian motion as in Section 2.2. Moreover, we put

$$\xi_{\ell,N} = |\sigma(t_\ell, X(t_\ell))| \tag{16}$$

with $t_\ell = \ell/k_N$, as previously.

LEMMA 2. *There exist measurable mappings $m_{\ell,N} : \mathbb{R}^{k_N+1} \to \mathbb{N}_0$ such that*

$$\liminf_{N \to \infty} (N \cdot E\,\|\rho_N(V_N)\,Y_N - X_N^\dagger\|_2^2) \geqslant \liminf_{N \to \infty} \left(\frac{N}{k_N^2 \pi^2} \cdot E\left(\sum_{\ell=0}^{k_N-1} \frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N) + 1} \right)\right) \tag{17}$$

and

$$\sum_{\ell=0}^{k_N-1} m_{\ell,N}(v) = \lceil E(v(X(0), W) \mid V_N = v \rceil. \tag{18}$$

If $\bar{X}_N \in \chi_N$ then (17) and (18) hold for constants $m_{\ell,N} \in \mathbb{N}_0$.

Proof. Clearly $\lambda(W) = \lambda(Y_N) + \lambda(\tilde{W}_{k_N})$ for $\lambda \in \Lambda$. Formally we may therefore consider $X_N^\dagger$ as a method that first evaluates W at the points t_ℓ and then applies bounded linear functionals to Y_N. The latter part requires N evaluations on the average. Clearly $\tilde{W}_{k_N}$ is measurable with respect to the σ-algebra generated by V_N, and Y_N consists of independent Brownian bridges (conditioned on $V_N = v$). We apply (12) to obtain

$$E(\|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2 \mid V_N = v) \geqslant \frac{1}{k_N^2 \pi^2} \cdot \sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(v)}{m_{\ell,N}(v)+1}$$

with measurable mappings $m_{\ell,N}$ that satisfy (18). Integration of this lower bound yields

$$E\, \|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2 \geqslant \frac{1}{k_N^2 \pi^2} \cdot E\left(\sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(V_N)}{m_{\ell,N}(V_N)+1} \right).$$

Due to (A) and (B),

$$E\, |\rho_{\ell,N}^2(V_N) - \xi_{\ell,N}^2| \leqslant c/k_N,$$

and therefore

$$E\|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2 \geqslant \frac{1}{k_N^2 \pi^2} \cdot E\left(\sum_{\ell=0}^{k_N-1} \frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1} \right) - c/k_N^2.$$

Using (13), we complete the proof in the general case.

Suppose that $\bar{X}_N \in \chi_N$. Formally, we may assume that $X_N^\dagger$ first evaluates W at the points t_ℓ and then applies fixed bounded linear functionals $\lambda_1, ..., \lambda_N \in \Lambda$ to Y_N. We conclude that

$$\begin{aligned} E(\|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2 \mid V_N = v) &\geqslant E\, \|\rho_N(v)\, Y_N - \rho_N(v)\, E(Y_N(\cdot) \mid \lambda_1, ..., \lambda_N)\|_2^2 \\ &\geqslant \frac{1}{k_N^2 \pi^2} \cdot \sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(v)}{m_{\ell,N}+1}, \end{aligned}$$

using (12) again. Now we proceed as above to complete the proof. ∎

We now separately analyze the classes χ^{**}, χ^{*}, and χ.

LEMMA 3. *If $\bar{X}_N \in \chi_N^{**}$ for every N then*

$$\liminf_{N\to\infty}\left(\frac{N}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\right)\geqslant (C^{**})^2.$$

Proof. Note that

$$E\left(\sum_{\ell=0}^{k_N-1} m_{\ell,N}(V_N)\right)\leqslant E(\nu(X(0),W))+1\leqslant N+1.$$

hence, by the Cauchy–Schwartz inequality,

$$\begin{aligned}
&\frac{N+k_N+1}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\\
&\quad\geqslant\frac{1}{k_N^2}\cdot\sum_{\ell=0}^{k_N-1}E(m_{\ell,N}(V_N)+1)\cdot\sum_{\ell=0}^{k_N-1}E\left(\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\\
&\quad\geqslant\frac{1}{k_N^2}\cdot\left(\sum_{\ell=0}^{k_N-1}\left(E(m_{\ell,N}(V_N)+1)\cdot E\left(\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\right)^{1/2}\right)^2\\
&\quad\geqslant\frac{1}{k_N^2}\cdot\left(\sum_{\ell=0}^{k_N-1}E(\xi_{\ell,N})\right)^2.
\end{aligned}$$

Now apply Fatou's Lemma and use (14) to obtain

$$\begin{aligned}
\liminf_{N\to\infty}\left(\frac{N}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\right)&\geqslant\left(E\left(\liminf_{N\to\infty}\left(\frac{1}{k_N}\cdot\sum_{\ell=0}^{k_N-1}\xi_{\ell,N}\right)\right)\right)^2\\
&=(C^{**})^2. \quad \blacksquare
\end{aligned}$$

LEMMA 4. *If $\bar{X}_N \in \chi_N^{*}$ for every N then*

$$\liminf_{N\to\infty}\left(\frac{N}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\right)\geqslant (C^{*})^2.$$

Proof. By assumption, $\nu(x,w)=n(\bar{X}_N)$ for every initial value x and every trajectory w. Thus

$$\begin{aligned}
(N+k_N)\cdot\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}&=\sum_{\ell=0}^{k_N-1}(m_{\ell,N}(V_N)+1)\cdot\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\\
&\geqslant\left(\sum_{\ell=0}^{k_N-1}\xi_{\ell,N}\right)^2,
\end{aligned}$$

which implies

$$\liminf_{N\to\infty}\left(\frac{N}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}(V_N)+1}\right)\right)\geqslant E\left(\liminf_{N\to\infty}\left(\frac{1}{k_N}\cdot\sum_{\ell=0}^{k_N-1}\xi_{\ell,N}\right)\right)^2$$
$$=(C^*)^2. \quad \blacksquare$$

LEMMA 5. *if $\bar{X}_N\in\chi_N$ for every N then*

$$\liminf_{N\to\infty}\left(\frac{N}{k_N^2}\cdot E\left(\sum_{\ell=0}^{k_N-1}\frac{\xi_{\ell,N}^2}{m_{\ell,N}+1}\right)\right)\geqslant C^2.$$

Proof. Clearly

$$(N+k_N)\cdot\sum_{\ell=0}^{k_N-1}E\left(\frac{\xi_{\ell,N}^2}{m_{\ell,N}+1}\right)=\sum_{\ell=0}^{k_N-1}(m_{\ell,N}+1)\cdot\sum_{\ell=0}^{k_N-1}\frac{E(\xi_{\ell,N}^2)}{m_{\ell,N}+1}$$
$$\geqslant\left(\sum_{\ell=0}^{k_N-1}(E(\xi_{\ell,N}^2))^{1/2}\right)^2,$$

and the statement follows. $\blacksquare$

This completes the proof of the lower bounds in Theorem 1.

3.4. *Proof of the Upper Bounds in Theorem* 1. The proof of the lower bounds motivates the following construction. Take a sequence of positive integers k_N satisfying (13) and (14). Consider the Milstein process $\check{X}_{k_N}$ and the piecewise linear interpolation $\tilde{W}_{k_N}$ of W that correspond to the discretization (4) with $k=k_N$. Let $f_{i,\ell,N}$ with $i\in\mathbb{N}$ and $\ell=0,\ldots,k_N-1$ denote the system of eigenfunctions associated with $Y_N=W-\tilde{W}_{k_N}$, see (10). Recall that $V_N=((X(0),W(t_1),\ldots,W(t_{k_N}))$.

We study methods of the form

$$\bar{X}_N(t)=\check{X}_{k_N}(t_\ell)+a(t_\ell,\check{X}_{k_N}(t_\ell))\cdot(t-t_\ell)$$
$$+\sigma(t_\ell,\check{X}_{k_N}(t_\ell))\cdot(\tilde{W}_{k_N}(t)-W(t_\ell))+X_N^\dagger(t)$$

for $t\in[t_\ell,t_{\ell+1}[$, where

$$X_N^\dagger(t)=\rho_{\ell,N}(V_N)\cdot\sum_{i=1}^{m_{\ell,N}(V_N)}\langle Y_N,f_{i,\ell,N}\rangle\cdot f_{i,\ell,N}(t).$$

Specific choices of measurable mappings $m_{\ell,N} : \mathbb{R}^{k_N+1} \to \mathbb{N}_0$ will be given below.

Note that $\bar{X}_N$ and $X_N^\dagger$ are related by (15). Hence Lemma 1 is applicable and it yields

$$\limsup_{N\to\infty} (N^{1/2} \cdot e(\bar{X}_N)) \leqslant \limsup_{N\to\infty} (N \cdot E \, \|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2)^{1/2}. \tag{19}$$

By (8) and (9)

$$\begin{aligned} E(\|\rho_N(V_N)\, Y_N - X_N^\dagger\|_2^2 \mid V_N = v) &= \frac{1}{k_N^2 \pi^2} \cdot \sum_{\ell=0}^{k_N-1} \rho_{\ell,N}^2(v) \sum_{i=m_{\ell,N}(v)+1} \frac{1}{i^2} \\ &\leqslant \frac{1}{k_N^2 \pi^2} \cdot \sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(v)}{m_{\ell,N}(v)}, \end{aligned} \tag{20}$$

where we use the convention $0/0 = 0$.

Now we determine the number of terms $m_{\ell,N}$ in the Karhunen–Loève approximation of Y_N on the subinterval $[t_\ell, t_{\ell+1}]$. Take $m_{\ell,N} = m_{\ell,N}^{**}$ with

$$m_{\ell,N}^{**}(v) = \lceil |\rho_{\ell,N}(v)| \cdot N/k_N \rceil$$

to define the method $\hat{X}_N^{**}$. Moreover, take $m_{\ell,N} = m_{\ell,N}^{*}$ with

$$m_{\ell,N}^{*}(v) = \begin{cases} \left\lceil (N - 2k_N) \cdot |\rho_{\ell,N}(v)| \Big/ \sum_{j=0}^{k_N-1} |\rho_{j,N}(v)| \right\rceil, & \text{if } \sum_{j=0}^{k_N-1} |\rho_{j,N}(v)| > 0 \\ \lceil (N - k_N)/k_N \rceil, & \text{otherwise.} \end{cases}$$

to define the method $\hat{X}_N^{*}$. Finally, take

$$m_{\ell,N} = \begin{cases} \left\lceil (N - 2k_N) \cdot \alpha_\ell \Big/ \sum_{j=0}^{k_N-1} \alpha_j \right\rceil, & \text{if } \sum_{j=0}^{k_N-1} \alpha_j > 0 \\ \lceil (N - k_N)/k_N \rceil, & \text{otherwise} \end{cases}$$

with

$$\alpha_\ell = (E(\sigma^2(t_\ell, X(t_\ell))))^{1/2}$$

to define the method $\hat{X}_N$, see (16).

Clearly $\hat{X}_N^{**} \in \chi^{**}$ and

$$\frac{N}{k_N} \sum_{\ell=0}^{k_N-1} \rho_{\ell,N}(v) \leqslant \sum_{\ell=0}^{k_N-1} m_{\ell,N}^{**}(v) \leqslant k_N + \frac{N}{k_N} \sum_{\ell=0}^{k_N-1} \rho_{\ell,N}(v).$$

Therefore

$$\lim_{N\to\infty} (1/N \cdot n(\hat{X}_N^{**})) = C^{**}, \tag{21}$$

see Hofmann *et al.* (2001, Lemma 7). Since

$$N - k_N \leqslant \sum_{\ell=0}^{k_N-1} m_{\ell,N}^{*}(v) + k_N \leqslant N,$$

we may assume $\hat{X}_N^{*} \in \chi_N^{*}$. Similarly, we may assume $\hat{X}_N \in \chi_N$.

We have

$$\frac{1}{k_N^2} \cdot \sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(v)}{m_{\ell,N}^{**}(v)} \leqslant \frac{1}{Nk_N} \cdot \sum_{\ell=0}^{k_N-1} \rho_{\ell,N}(v),$$

so that

$$\limsup_{N\to\infty} (N \cdot e^2(\hat{X}_N^{**})) \leqslant C^{**}/\pi^2 \tag{22}$$

follows from (19) and (20). Combining (21) and (22) we obtain the upper bound in Theorem 1(i).

Next, observe that

$$\frac{1}{k_N^2} \cdot \sum_{\ell=0}^{k_N-1} \frac{\rho_{\ell,N}^2(v)}{m_{\ell,N}^{*}(v)} \leqslant \frac{1}{N-2k_N} \cdot \Big(\frac{1}{k_N} \cdot \sum_{\ell=0}^{k_N-1} |\rho_{\ell,N}|(v) \Big)^2.$$

Using (19) and (20) we get

$$\limsup_{N\to\infty} (N \cdot e^2(\hat{X}_N^{*})) \leqslant (C^{*}/\pi)^2,$$

which yields the upper bound in Theorem 1(ii).

The upper bound in Theorem 1(iii) immediately follows from (19), (20), and Hofmann *et al.* (2001, Lemma 10).

3.5. *The Wagner–Platen Process.* For the analysis of the Wagner–Platen scheme we replace the Milstein process by the Wagner–Platen process $\check{X}_n$. The latter is defined by the equidistant discretization (4) together with $\check{X}_n(0) = X(0)$ and

$$\check{X}_n(t) = \check{X}_n(t_\ell) + \sum_{j=1}^{7} g_j(t_\ell, \check{X}_n(t_\ell)) \cdot Z_{j,\ell}^{(n)}(t)$$

for $t \in \,]\, t_\ell, t_{\ell+1}]$. Obviously $\check{X}_n(t_\ell) = \hat{X}_n^{\mathrm{equi}}(t_\ell)$ for $\ell = 0, \ldots, n$.

From Kloeden and Platen (1995, Theorem 10.6.3) we know that

$$\sup_{t \in [0,1]} E(X(t) - \check{X}_n(t))^2 \leqslant c/n^3, \tag{23}$$

provided (A′) and (B) are satisfied. In (23) and in the remainder of the paper c denotes unspecified positive constants, which only depend on the constant K from (A′) as well as on $g_1(0,0), \ldots, g_7(0,0)$ and $E(X(0))^4$.

3.6. *Proof of the Upper Bound in Proposition* 1. From (23) we conclude that

$$e(\hat{X}_n^{\text{equi}}) \leqslant (E\,\|\check{X}_n - \hat{X}_n^{\text{equi}}\|_2^2)^{1/2} + c/n^{3/2}. \tag{24}$$

Therefore it remains to show that

$$\limsup_{n \to \infty} n^{1/2} \cdot (E\,\|\check{X}_n - \hat{X}_n^{\text{equi}}\|_2^2)^{1/2} \leqslant C^{\text{equi}}/\sqrt{6}\,.$$

Put

$$S_\ell = (t_\ell, \hat{X}_n^{\text{equi}}(t_\ell)).$$

Let $t \in [t_\ell, t_{\ell+1}]$. Then we have

$$\begin{aligned}
\check{X}_n(t) - \hat{X}_n^{\text{equi}}(t) = {} & g_2(S_\ell) \cdot (W(t) - \tilde{W}_n(t)) \\
& + g_3(S_\ell) \cdot ((W(t) - W(t_\ell))^2 - n \cdot (t - t_\ell) \cdot (W(t_{\ell+1}) - W(t_\ell))^2) \\
& + g_4(S_\ell) \cdot (t - t_\ell) \cdot (W(t) - W(t_{\ell+1})) \\
& + g_5(S_\ell) \cdot ((W(t) - W(t_\ell))^3 - n \cdot (t - t_\ell) \cdot (W(t_{\ell+1}) - W(t_\ell))^3) \\
& + g_6(S_\ell) \cdot ((t - t_\ell)^2 - (t - t_\ell)/n) \\
& + g_7(S_\ell) \left(\int_{t_\ell}^{t} (W(s) - W(t_\ell))\, ds \right. \\
& \left. - n \cdot (t - t_\ell) \int_{t_\ell}^{t_{\ell+1}} (W(s) - W(t_\ell))\, ds \right).
\end{aligned}$$

In the sequel we use the linear growth

$$|g_j(t, x)| \leqslant c \cdot (1 + |x|).$$

of the coefficient functions. Note that (23), (A′), and (3) imply the boundedness of $E(\hat{X}_n^{\mathrm{equi}}(t_\ell))^2$ and that $W(t)-W(t_\ell)$ and S_ℓ are independent if $t \geqslant t_\ell$. We obtain

$$E(g_3(S_\ell)\cdot((W(t)-W(t_\ell))^2-n\cdot(t-t_\ell)\cdot(W(t_{\ell+1})-W(t_\ell))^2))^2 \leqslant c/n^2$$

and

$$E(g_4(S_\ell)\cdot(t-t_\ell)\cdot(W(t)-W(t_{\ell+1})))^2 \leqslant c/n^3.$$

Furthermore,

$$E(g_5(S_\ell)\cdot((W(t)-W(t_\ell))^3-n\cdot(t-t_\ell)\cdot(W(t_{\ell+1})-W(t_\ell))^3))^2 \leqslant c/n^3$$

and

$$E(g_6(S_\ell)\cdot((t-t_\ell)^2-(t-t_\ell)/n))^2 \leqslant c/n^4.$$

Clearly

$$E\left(\int_{t_\ell}^{t}(W(s)-W(t_\ell))\,ds\right)^2 \leqslant c/n^3.$$

Therefore,

$$E\left(g_7(S_\ell)\cdot\left(\int_{t_\ell}^{t}(W(s)-W(t_\ell))\,ds-n\cdot(t-t_\ell)\int_{t_\ell}^{t_{\ell+1}}(W(s)-W(t_\ell))\,ds\right)\right)^2 \leqslant c/n^3.$$

Summarizing we get

$$(E\,\|\check{X}_n-\hat{X}_n^{\mathrm{equi}}\|_2^2)^{1/2} \leqslant \left(\int_0^1 E(\sigma^2(S_\ell)\cdot(W(t)-\tilde{W}_n(t))^2)\,dt\right)^{1/2}+c/n. \qquad (25)$$

Hence

$$\begin{aligned}&\limsup_{n\to\infty} n^{1/2}\cdot(E\,\|\check{X}_n-\hat{X}_n^{\mathrm{equi}}\|_2^2)^{1/2}\\ &\quad\leqslant \limsup_{n\to\infty} n^{1/2}\cdot\left(\sum_{\ell=0}^{n-1} E(\sigma^2(S_\ell))\int_{t_\ell}^{t_{\ell+1}} E(W(t)-\tilde{W}_n(t))^2\,dt\right)^{1/2}.\end{aligned}$$

Using $\int_{t_\ell}^{t_{\ell+1}} E(W(t)-\tilde{W}_n(t))^2\,dt = 1/6n^2$ and (A′), (3), and (23), we conclude that the right-hand side is bounded from above by $C^{\mathrm{equi}}/\sqrt{6}$.

3.7. *Proof of the Lower Bound in Proposition* 1. In addition to (24) and (25) we also have

$$e(\hat{X}_n^{\mathrm{equi}}) \geqslant (E\,\|\check{X}_n - \hat{X}_n^{\mathrm{equi}}\|_2^2)^{1/2} - c/n^{3/2}$$

and

$$(E\,\|\check{X}_n - \hat{X}_n^{\mathrm{equi}}\|_2^2)^{1/2} \geqslant \left(\int_0^1 E(\sigma^2(S_\ell)\cdot(W(t)-\tilde{W}(t))^2)\,dt \right)^{1/2} - c/n.$$

Thus

$$\liminf_{n\to\infty} n^{1/2}\cdot e(\hat{X}_n^{\mathrm{equi}}) \geqslant \liminf_{n\to\infty} n^{1/2}\cdot(E\,\|\check{X}_n - \hat{X}_n^{\mathrm{equi}}\|_2^2)^{1/2} \geqslant C^{\mathrm{equi}}/\sqrt{6}\,.$$

ACKNOWLEDGMENT

The third author (K.R.) was partially supported by the Alexander von Humboldt Foundation.

REFERENCES

N. Hofmann, T. Müller-Gronbach, and K. Ritter, The optimal discretization of stochastic differential equations, *J. Complexity* **17** (2001), 117–153.

P. Kloeden and E. Platen, "Numerical Solution of Stochastic Differential Equations," Springer-Verlag, Berlin, 1995.

D. Lee, Approximation of linear operators on a Wiener space, *Rocky Mountain J. Math.* **16** (1986), 641–659.

G. N. Milstein, Approximate integration of stochastic differential equations, *Theory Probab. Appl.* **19** (1974), 557–562.

T. Müller-Gronbach, Best rates of convergence for strong approximation of an sde at a single point, in preparation, 2001.

A. Papageorgiou and G. W. Wasilkowski, On the average complexity of multivariate problems, *J. Complexity* **6** (1990), 1–23.

K. Ritter, "Average-Case Analysis of Numerical problems," Lecture Notes in Mathematics, Vol. 1733, Springer-Verlag, Berlin, 2000.

J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, "Information-Based Complexity," Academic Press, New York, 1988.

W. Wagner and E. Platen, Approximation of Ito integral equations, preprint, ZIMM, Akad. Wiss. DDR, Berlin, 1978.

G. W. Wasilkowski, On adaptive information with varying cardinality for linear problems with elliptically contoured measures, *J. Complexity* **5** (1989), 363–368.

G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for weighted approximation, preprint, 2001.